USE OF MOVING AVERAGES AND INTERPOLATION TO ESTIMATE MEDIAN-EFFECTIVE DOSE

I. Fundamental Formulas, Estimation of Error, and Relation to Other Methods[1]

WILLIAM R. THOMPSON

Division of Laboratories and Research, New York State Department of Health, Albany

Contents

Estimation of the potency of a given reagent by means of its action upon living matter is the purpose of biologic assay. Test designs and methods of evaluation of results have been developed extensively through collaboration of biologists

[1] Presented at a meeting of the Eastern New York Branch of the Society of American Bacteriologists, Albany, N. Y., November 30, 1945; Abstract in Annual Report of the Division of Laboratories and Research, New York State Department of Health, 1945, p. 20. A condensed version appears as part of a chapter on biologic assay in the third edition of the Standard Methods of this laboratory (5a).

and statisticians. Methods of estimation of median-effective dose[2] (M) have been a critical feature of many systems of *quantal* assay, where the response of test-individuals is *all or none*. The present article describes a method of moving-average interpolation to estimate M from the assay data. Section headings and table of contents may serve as a guide to information essential to its actual use or to an understanding of its fundamental character and relation to other methods. Fine print has been used where discussions serve more as amplification or aid to a careful study than as development of the main themes. On account of the relative simplicity of assumptions and calculations involved, the present method may be preferred unless the use of some other method in a given situation can be justified by improved precision or by permitted technical economies. It is in this sense that moving-average interpolation is suggested as a *basic method*, not in the sense that it is more elegant aside from this dual simplicity. Attention is directed also to judgment of possible influences of choice of test-plans upon precision.

INTRODUCTION

In biologic assay based on quantal, *all or none*, response it has been found widely advantageous to use a sequence of doses (D_i) in geometric progression, for $i = 1, \cdots, h$, each administered to a given number n_i of subjects (such as animals or eggs); and it seems preferable to plan to have the same number of subjects in each case, $n_i = n$. In each case the number r_i that respond critically may be tabulated, against the corresponding dose D_i in ascending order of magnitude. In any case the n_i subjects and results of administration of the dose D_i are considered as a sample from a hypothetical universe of all possible individuals of the sort (respectively subjects or results) presumably obtainable by

[2] The writer's attention has been called to the unfortunate fact that M has been widely used with a different meaning elsewhere in the literature on bioassay, as by Gaddum (22) to represent the logarithm of the ratio of two potencies. To facilitate the present discussion, in describing the logistic relation, we here employ M, D, p, and G respectively where Thompson and Maltaner (7) and also Kent *et al.* (9) used K, x, y, and h or $1/n$, and where von Krogh (2) used k, x, y, and $1/n$.

It is much more convenient here to use M to denote the median-effective dose than alternative notations such as ED_{50}. Moreover, this is in accord with the notation employed in the chapter on biologic assay in the Standard Methods (5) of this laboratory which also employs confidence ranges for the median M with the same notation used in their original development (36) in a form yielding certain exact probability statements, *specific confidence* that any given percentile of essentially any universe lies below the kth observation in ascending order from a random sample of aggregate number n. The article has been discussed in a review by C. C. Craig (39a). Use of any other symbol here instead of M would seem to place an unwarranted burden on cross-reference. Furthermore, M for the median occurs in many statistical texts, e.g., Rietz (29), Camp (40a), and Rider (28); others employ more cumbersome symbols, such as Mi, Md, Med, or Q_2.

It is suggested that an *italic* or *script* L or λ might be used for the logarithm of the ratio of two potencies instead of Gaddum's M, or a *script* M be used for the median or median-effective dose instead of the present M, if economy demands single-lettered representation of each in the same discussion. The second alternative seems less attractive since the symbols do not suggest their meaning.

more extensive experience. Such a universe is conveniently called the sampled *population*, usually considered as having an infinite aggregate number. Let $p_i \equiv r_i/n_i$, and $\bar{p}_i$ be the hypothetically unknown *true* probability of critical response to the dose D_i for the sampled population. Then, the usual immediate objective in such a bioassay is to estimate the *median-effective* dose M; defined as a dose such that, if $D_i = M$, then $\bar{p}_i = 0.5$ (i.e., M is the dose that should be *expected* to yield a 50-per-cent response; for example, the LD_{50}).[2]

For convenience in discussion, let the hypothetical curve of points $(\log D, \bar{p})$ be called the *fundamental curve*; and consider its hypothetical graph to have $\log D$ as abscissa and $\bar{p}$, the corresponding *true* probability of critical response, as ordinate. If some function of $\bar{p}$ is used instead as ordinate, say $T_{(\bar{p})}$, let the corresponding points and curve be regarded as *transformed* from the fundamental coordinate system $(\log D, \bar{p})$; consider experimental points $(\log D_i, p_i)$ as likewise *transformed* to points $(\log D_i, T_{(p_i)})$.

The usual purpose of such a transformation is to straighten the fundamental curve so that a straight line may be fitted to the transformed points and M be estimated by M' from the point $(\log M', T_{(0.5)})$ on the fitted line, corresponding to $p = 0.5$. The possibility of a gain in power of estimation by such methods is obvious. However, there is imminent danger that tendencies toward biased or erratic estimates may be induced by mistaken assumptions about the form of the fundamental curve or by the technics used for curve fitting. Even for no other purpose than to serve as a basis of comparison, it would be advantageous to have available an objective, unbiased method of estimating M, free from assumption as to the precise type of fundamental curve involved, but capable of taking into account more of the data than that from successive doses where p_i and p_{i+1} straddle 0.5. It would seem wise to choose such a method as *basic* (i.e., used as a basis of comparison of other methods under consideration), at least in situations where there is so little information about the form of the fundamental curve as to make any assumption about it hazardous. If calculations involved in the basic method were much simpler than those required in some rival method, then the latter should bear the burden of proof that any expectation of improvement it could offer would be worth the added effort.

The purposes of this communication are: to present such a basic method with formulas for simple direct calculation of $\log m$ as an estimate of $\log M$, to show (in Appendix) how the variance and standard deviation of $\log m$ may be estimated from an individual assay experience in default of a broader basis for their estimation or, for purposes of comparison, to examine some of the characteristics of the present and other methods that may influence a choice among them in view of other conditions governing a given assay system, and to illustrate use of the present method in comparison with certain others on a body of data that has been presented independently for a like purpose. The proposed method is founded upon a well-known system of graduation by so-called *moving averages* followed by interpolation, all effected by use of a relatively simple formula in conjunction with a simple test of data under consideration to indicate whether the result will be an interpolation (as desired) or an extrapolation, if a given

range is used as a basis for the calculation. A preliminary discussion of other methods is given below as well as an outline of test plans and sampling technics to furnish a background and foundation for development of formulas and discussion of the relative precision to be expected. Random and stratified random sampling technics are outlined, that are applicable not only to quantal assays but to other assays, which may be called *gradational.*

In use of methods of estimating M by fitting a curve of given type, the moderate assumption is usually implied that with a suitable limitation of range $(0 < \alpha \leqq \bar{p} \leqq \beta < 1)$ the given curve-type so nearly approximates the form of the fundamental curve as to make a satisfactory substitute under the circumstances. Two curve types, the *logistic* and the *integrated normal,* appear to have been most favored. Winsor (1) has shown that either may be fitted to the other so well over the ranges usually employed in bioassay that it would ordinarily be difficult to discriminate between them on the basis of goodness of fit to experimental data and usefulness as a means of estimating M.

Complement-Fixation Assay Using Logistic Function. von Krogh (2) used the logistic function and a transformation, $T_{(p)} = \log \frac{p}{1-p}$, to the coordinate system $\left(\log D, \log \frac{p}{1-p}\right)$ to represent the curve of hemolysis of red blood cells, finding that experimental data yielded points in the transformed system that lay approximately on a straight line,[2]

$$\log D = \log M + G \cdot \log [p/(1-p)]. \tag{i}$$

Obviously, this relation is independent of the base of the logarithms; for theoretical discussions the base e is more convenient, but for calculations the base 10 may be used. In any case we may convert from one form to the other by the well-known identity, $\log_e x = \log_{10} x \cdot \log_e 10 \cong 2.30259 \log_{10} x$. Relation (i) has been used as *fundamental curve* in evaluation of bioassays (3–10, 38) that employ such hemolysis as indicator in the titration of various antigen-antibody aggregates (or either component in the presence of a maximally reactive amount of the other) by means of their complement-fixing ability under prescribed conditions—the unfixed residual complement acting as the hemolytic agent on the previously sensitized cells. The total number (n) of cells used in each case is large (roughly estimated by Elizabeth and Frank Maltaner as about 125 million in their work and by Kent as about 100 million in his work with Bukantz and Rein (9)); p is estimated colorimetrically, and a restricted range for p is specified as admissible for use in evaluations.

It might at first glance be supposed that the random sampling errors in p would be a dominant influence and that therefore the linear relation (i) should be fitted so as to minimize a weighted sum of squared deviations of $\log [p/(1-p)]$ from the fitted line. However, by well-known relations the standard deviation of p (with all else the same in random sampling) would be $\sqrt{\bar{p}(1-\bar{p})/n}$ and twice this cannot exceed $1/\sqrt{n} \leqq 0.0001$ for $n \geqq 100$ million. Accordingly, this source of error would appear to have a negligible rather than a dominant influence (with deviations of p from $\bar{p}$ numerically less than 0.0001 in nineteen

out of twenty trials). Indeed, p is estimated in most routine tests by visual comparison with standards to the nearest 0.05 or, as in the work of Kent *et al.* (9), photoelectrically to about the nearest 0.01. Errors in measuring reagents may well exert a greater influence on the results (12c). Thus the procedure they employed, minimizing the sum of squared deviations in log D, should not be rejected offhand on this basis. The writer has recommended to Mr. Kent, the Maltaners, and Doctor Rice, well-known methods (10a) of simplifying the calculations and of relatively simple extension to the case where a family of k parallel regression lines are fitted to k sets of data such as those obtained in the experiments of Kent *et al.* (9). However, it is interesting to calculate the regression lines fitted individually to each set to see what variation in slope occurs, and likewise to compare the results of fitting of straight lines by inspection from a graph in the transformed coordinates (the technic actually used in most of the earlier work).

Now, the purpose of Kent, Bukantz, and Rein was essentially to develop more convenient and precise graphs to be used subsequently for estimation of M from the value of p obtained in a single hemolysis test with a given dose of complement D in a domain where the slope G of the regression lines (i) could be considered approximately constant. This situation is realized in complement titration (where antigen, serum, and antibody are absent) and in the case of the system investigated by the Maltaners (6) in which egg albumen was used as antigen. Thus relation (i) with G evaluated in previous experience could be used directly or in a graph or table to give M from the amount of complement D used and the degree of hemolysis p that was observed to result in the conventional test. Kent *et al.* rejected results where p lay outside the interval (0.2, 0.8) but studied relative errors both within and without this interval in formal application of their method. With restriction of p thus, it appears (in their table 6) that for estimation of M the standard deviations were about 2 per cent or less. Thus far we may suppose simply that M and D are expressed in the same arbitrary unit system. However, it is convenient for many purposes to take the arbitrary unit so that $M = 1$ in the case of complement titration. The volume of a given complement preparation that corresponds to one arbitrary unit is estimated by a preliminary or simultaneous test or both.

With other antigen-antibody systems such as those designed to study reactivity of tuberculous, syphilitic, gonococcal, pneumococcal, or viral antisera with homologous antigens (4, 5, 8, 10) an additional difficulty appeared. Thus it was found with the syphilitic system that with a given amount of serum and the optimal amount of antigen a relation of form (i) was approximated in all titrations with various doses D but that the slopes G of the regression lines varied greatly, and approximately systematically as a function of M. The writer's first contact with these investigations was when Dr. Frank Maltaner asked for help in solving the problem: given a smooth graduation curve to represent the dependence of G on M, and that for a given M relation (i) holds as stated above, how can the unknown M be estimated from the degree of hemolysis p observed in a single test employing a given amount of complement D. A graphic method for solving this problem was developed (7) and used to construct graphs and tables for each of the systems investigated, so that the appropriate graph or table could be entered with the given values of p and D to obtain the

required value of M directly. This method may be designated briefly as the *modulation method* with the intention of suggesting that it takes into account the gradual variation in G.

The original article (7) shows how ratio tables may be constructed so that data from tests with two reaction mixtures (e.g., one being a *complement* or a *serum control*) and having D and p values respectively (D_1, p_1) and (D_2, p_2) may be applied to an appropriate table (for D_2, D_1) making a cross entry with p_2 and p_1 to find directly the ratio $I = M_2/M_1$. A number of tables so constructed were published (4) prior to the method itself. Their form was made to accord with that customary in reporting routine results with a given specimen as the ratio I with M_2 from a test with optimal amount of antigen and M_1 without antigen. Possibly this custom is only a temporary expedient pending further investigation; but it has been used because it was believed (3, 4) that to a great extent the anticomplementary activity exhibited by certain sera would have a roughly proportional effect on complement whether antigen was present or not. However, the issue is not clear (38).

The egg albumen system (as has been mentioned) gave a constant G, and so did the pneumococcal system at 37° but not at 3° to 6°. In that case and in all other systems investigated G was found inconstant but largely dependent on M, though differently in each, but permitting application of the modulation method. Whether or not estimates of M obtained on this basis furnish a useful measure of something must rest upon demonstration with each antibody-antigen system. However, all of those investigated yielded in this way values for M that were found approximately linearly related under given conditions to the relative amount of either antigen or antibody in the reaction mixture if the other agent was present in an approximately maximally-reactive amount. This furnished a justification for use of the estimation method with these systems. Recent investigations by Rice with the pneumococcal (8, 38) and vaccinia viral (10) systems afford examples. For some time attention has been focused upon improvement of reagents as in use of the pure substance cardiolipin in tests for syphilis (41, 42).

Some estimates of reproducibility of the ratio I in routine tests have been made by constructing confidence ranges for the relative numerical discrepancy $\delta = 2(I_1 - I_2)/(I_1 + I_2)$, where $I_1 \geqq I_2$ denote replicate values of I obtained in tests with the same specimen (as in serum tests for syphilis), the replicates being either simultaneous with the same reagents or obtained on different days with different reagent preparations. The statistical analysis was based on methods described elsewhere (36, 46, 47) to obtain from sample experiences values of δ_P such that the probability of encountering a value of $\delta > \delta_P$ in any given future test under the same conditions was P. Estimates of $\delta_{0.5}$ and $\delta_{0.1}$ were obtained for I in different ranges in the test for syphilis and found to be nearly the same. The estimate from a pooled experience with 268 pairs of nonsimultaneous observations gave $\delta_{0.5} \cong 8$ per cent and $\delta_{0.1} \cong 25$ per cent. For simultaneous replicates a sample of 80 pairs gave $\delta_{0.5} \cong 5$ per cent and $\delta_{0.1} \cong 16$ per cent.

Some Characteristics of Other Assay Methods. A review of other literature on biologic assay has been presented by Bliss and Cattell (11) with a discussion of some of the fundamental concepts, which are also the subject of an interesting article by Irwin (12a) and discussion that followed its presentation (12b);

especially noteworthy is an appended contribution by Neyman (12c). For a given value of p, Bliss defined the *probit* as five plus the equivalent *normal deviation* (with unit standard deviation); the transformation, $T_{(p)}$ = probit p, converts the integrated normal curve to a straight line that passes through the transformed point (log M, 5), as probit 0.5 = 5. This, or the corresponding normal deviate transformation, has been applied extensively to assay systems where n is large (11, 13), and has been modified (14–16) for use where n is small (in which case $p_i = 0$ or 1 is not uncommonly encountered). However, in the latter case the calculations required in estimation of M are somewhat difficult, involving successive approximations with tentative regression lines fitted by a method of *maximum likelihood*. The logistic function has been applied instead by Wilson and Worcester (17, 18) using maximum likelihood, and by Berkson (19, 20) using a method of weighted least squares for curve fitting. In terms of the natural (base e) logarithms Berkson (19) defines logit $p = \log[(1 - p)/p] = -\log[p/(1 - p)]$, transforming to logits prior to the curve fitting (the negative of von Krogh's transformation (2)). He gives some comparisons between uses of the logistic and of the integrated normal curve in applications to the same data in a variety of situations, the comparisons uniformly in favor of the logistic. Occurrence of values of p_i equal to zero or one is a source of difficulty in fitting either the integrated normal or the logistic curve to data, as the transformation to either probit or logit gives infinite values in these cases. As a result, a special treatment is employed in the form of an adjustment of such data.

Some attempts have been made to develop methods of estimating M that avoid the definite assumption as to the fundamental curve form and attendant difficulties of curve fitting, but emphasis has been placed mostly upon facility in calculation; most prominent have been the Kärber (21–24) and the Reed-Muench (25) methods. Apparently, the most generally applicable method is the obvious device of simple interpolation between successive values of p_i that happen to straddle 0.5. This is included as a special case ($K = 1$) in the moving-average method to be presented. However, this simple procedure makes no use of other assay data except in scanning, and thus in practice appears inefficient. Irwin and Cheeseman (23, 24) have used data of Mrs. Joyce Wilson and Professor Topley, given in the present table 1, to compare the method of Bliss (14, 15) with that of Kärber, which they call respectively "the exact" and "the approximate" method. However, Kärber's method is open to several objections; it may be shown to lead, even under the most ideal conditions, to the paradoxical conclusion that the approximate median-effective dose of any toxin is *none whatever*, unless it is assumed that subjects given very small doses of toxin (or none) would bear charmed lives during the experimental interval of observation. In general, it seems impossible to avoid some such vitiating feature with any suggested method that does not include an objectively applied restriction of the range of values of D_i that are allowed to influence the estimation. Although any good method may be made ineffective by an inadequate experimental plan, too meager data, or extraordinarily erratic experimental results, the procedure should otherwise lead to unbiased unequivocal estimates. Reed and Muench

(25) warned against effects of failure to restrict range properly, but their method has been abused regardless of the warning. Furthermore, it can be demonstrated that the Reed-Muench method, even with the absolute antithesis of erratic data (a uniform trend), has an unfortunate predilection to yield biased or equivocal estimates. Accordingly, neither the Kärber nor the Reed-Muench method meets the basic need for an objective, unbiased method of estimating M, free from assumption as to the precise type of fundamental curve involved, but capable of taking into account more than the data from successive doses where corresponding values of p_i straddle 0.5. Not only is this need met by the proposed moving-average method; but in simplicity of calculations it appears to be best, except that Kärber's method may be regarded as a degenerate form of the moving-average method and, accordingly, a much simpler calculation could be used instead of that traditionally employed (21, 23). Further discussion of these relations will follow a consideration of experimental plans in a quantal assay and adoption of some conventions as an aid to discussion.

EXPERIMENTAL PLANS

Fundamental Curve and Sense Convention. In any quantal assay system to be considered, it is assumed that the sampled population may be regarded as infinite and that $\bar{p}$, the *true* probability of critical response to the dose D in the sampled population, is either an *increasing* or a *decreasing* function of D within a convenient range of dosage including the median-effective dose M. Unless otherwise stated, no other specific knowledge of the form of this function is assumed. The fundamental curve is treated as if existing though unknown, formed from the hypothetical points (log D, $\bar{p}$) with log D as abscissa and $\bar{p}$ as ordinate. Strictly speaking, the curve is only a fiction, convenient to use in discussion of the population sampled in any given assay test or set of tests such that the subjects may be considered as drawn from the same population.

Even in use of the same colony of potential subjects for test, it may not be advisable to make such an assumption if a considerable time lapse or other possibly disturbing circumstance intervenes between tests; perhaps the fundamental curve should be considered as altered. Age increase alone may alter potential susceptibility of individuals, and thus the curve may be changed. To say that the fundamental curve is different implies, of course, that we are dealing with a different *sampled population* in the idealized sense intended. Such possible changes in curve form through seasonal variations, generation peculiarities, and secular trends might make it practically impossible ever to have enough information about the fundamental curve to be useful in curve fitting.

The critical response may be defined in any assay arbitrarily as either the occurrence or non-occurrence of a specified result to individuals (for example, death or survival), obviously without affecting the value of the median-effective dose M. Accordingly, to fix the ideas, a *sense convention* will be adopted: it will be assumed that the choice is so made that $\bar{p}$ is an increasing function of log D (necessarily then of D also) at least within a restricted range as mentioned above. It is easily demonstrated as a check on essential formulas (and tests for interpolation) that they hold equally if q is substituted for p and vice versa and

s_i for r_i and vice versa, where $q = 1 - p$ and $s_i = n_i - r_i$. This shows the sense convention to be merely a convenience in discussion.

Preliminary Conditions and Organization of Tests. In any assay system a suitable choice of species, and conditions of breeding, maintenance, and preparation of *subjects* for tests are of prime importance, as well as restrictions that may be made with regard to certain attributes such as age, weight, apparent conditon, and sex; the possibilities of attainment of greater assay precision by modification of such characteristics of a test plan are discussed by Bliss and Cattell (11a). For any given set of tests wherein conditions are to be made as comparable as possible aside from such specifications, it is best to use samples in each case drawn in an unbiased manner from a pool of prospective test-subjects that have been selected as suitable in accord with the specifications. Suppose that there are H ultimate categories (cases), possibly testing various reagent preparations in a given test plan; and suppose that n test-subjects are required in each case. Then the pool should contain at least $n \cdot H$ subjects; and H cells should be available, numbered $1, \cdots, H$, to correspond to the H cases so that n subjects may be placed conveniently in each cell as they are assigned by the sampling procedure. Two systems of unbiased sampling are given below, based on principles and terminology discussed by Neyman (26a).

Random Sampling (Unrestricted). An aggregate of $n \cdot H$ tags, numbered n each $1, \cdots, H$, but otherwise alike, are placed in a bowl, shuffled, and indiscriminately withdrawn one at a time without replacement as an individual is taken from the pool and assigned to the correspondingly numbered cell. The term *random sampling* implies unrestricted random sampling unless otherwise indicated.

Stratified Random Sampling. Many systems of stratified random sampling are possible; for example, the H subjects in each stratum may be selected in ascending order of weight or other attribute from the pool, or each stratum may contain individuals alike in sex, or characterized by certain combined attributes. Instead of the preceding technic, a pool of $n \cdot H$ subjects may be initially subdivided into n arbitrary subclasses, called *strata.* Only one tag for each number $(1, \cdots, H)$ is placed in the bowl and these are withdrawn at random, as above, without replacement as the subjects of any one stratum are assigned to cells corresponding. After the random sampling of the first stratum, all tags are returned to the bowl and the sampling process is repeated successively with the other strata. In the end, each cell has n subjects, one taken at random from each stratum. This is called a *stratified random sample.* If the strata are simply formed by the order of withdrawal of subjects from a general pool of $n \cdot H$ or more, obviously the strata need not be preformed. This may be called *simply stratified sampling.*

Difficulties in probability calculations, sometimes apparently insurmountable, are introduced if an unrestricted random sample is not employed, but a stratified random sample may be preferable in some situations. Thus in dealing with a pool of animals, some of which have greater ability to elude the sampler so that other less agile subjects are sooner picked for distribution in the cells, the simply stratified sampling technic would assure having

one of the first H in each cell and likewise one of each successive set of H animals so taken. Such agility, if related to variation in resistance to the agent one way or another, or suspected of having some relation, might throw the weight of judgment in favor of stratification. In general, the resultant precision should be as good or better if stratified rather than unrestricted random sampling is used. Consideration of what forms of stratification should be tried in any given situation, and investigation of the relative merits of rival forms is a possibly important aspect of any problem of bioassay that is usually ignored entirely. Even worse, a bias may be carelessly or inadvertently introduced in the sampling technic; for example, by the erroneous procedure of assigning the first n animals captured to the first cell, etc., in a fond belief that a random sampling would result.

Simultaneous Comparison of Standard and Unknown. The usual procedure in either quantal or gradational assay is to use a standard reagent to furnish a relative value for the unknown. The standard usually should be a preparation of the same active agent, such as a standard antipneumococcus serum of the same type as another preparation that is to be tested. The usually great importance of simultaneous comparisons of this sort deserves the emphasis that it has received in the literature (5a, 11, 12). It is rare that the biologic system on which the reagent acts may itself be used satisfactorily as a standard. However, a striking example of this is afforded by a quantal assay employing mortality of the eggs of *Drosophila melanogaster* as a biologic indicator of x-ray dosage (43). An interesting contrast is furnished by the gradational assay system employing prolongation of the larval stage of the Drosophila by x-ray irradiation (31–33), which system affords an illustration (5a) of several types of difficulty to be encountered in either quantal or gradational bioassay.

In the preceding scheme of sampling it may be supposed that usually some of the H ultimate categories (cases), corresponding to cells (cages) each containing n subjects, are to be reserved for use of the standard preparation. In the case of the data to be discussed below (given in table 1), there are ten replicate assays run simultaneously with the same agent; but these are treated separately in estimation of M, as if they were all made on different preparations (or preparations not *known* to be the same). Any one of these assays (A′ to K′) could be considered as the *standard* preparation; but we need not be concerned primarily with comparison of *standard* and *unknown* for the purposes of the following discussion, which is directed toward consideration of variation in such estimates of M employing a given method, and cross-comparison of estimates obtained by different methods from the same data.

The purpose is similar to that of Irwin and Cheeseman in employing the same data (23, 24). It seems obvious that such a purpose of comparison of methods applied to replicate assays would not be served by any attempt to pool information from all the assays (A′ to K′) in order to estimate some useful characteristics of the fundamental curve that might then be applied to the individual estimates. However, it is hoped that the present use of a particular test plan and basis of comparison will not appear as a detraction of other methods which may permit direct comparisons of relative potency (possibly without reference to median-effective dose) or abbreviation of tests as in the complement-fixation systems discussed above and in other assay systems (45).

DEVELOPMENT OF FORMULAS FOR INTERPOLATION FROM MOVING AVERAGES

Now, consider certain formal relations that lead to a development of formulas that yield by simple calculations an estimate m of the median-effective dose M. Suppose we have data in the previously indicated form of associated values of D_i and r_i for a given number n_i of subjects in each case, usually with $n_i = n$ resembling any column A′ to K′ in table 1. Assume that these have been obtained in accord with a suitable test plan with a set of h doses (D_i) wherein $D_j = R \cdot D_{j-1}$ and R is a constant greater than one ($i = 1, \cdots, h; j = 2, \cdots, h$).

TABLE 1

Deaths (r_i) in ten differently labeled samples of mice injected with given doses (D_i) of the same toxin preparation. $L_i = \log D_i$*

i	L_i	D_i	A′	B′	C′	D′	E′	F′	G′	H′	J′	K′
		mg										
1	$\bar{2}.7959$	0.0625	1	1	0	2	0	0	0	1	0	1
2	$\bar{1}.0969$	0.125	2	2	0	0	0	0	0	3	0	0
3	$\bar{1}.3979$	0.25	3	1	5	5	3	2	4	2	3	5
4	$\bar{1}.6990$	0.5	5	5	4	5	4	1	3	5	3	4
5	0.0000	1.0	5	4	4	5	5	5	5	5*	2	5
6	0.3010	2.0	5	5	5	5	5	5	5	5	5	5
7	0.6021	4.0	5	5	5	5	5	5	5	5	5	5

* Note: Data of Wilson and Topley used by Irwin and Cheeseman (23). In another article Irwin and Cheeseman (24) give the same table except for an apparent typographic error giving 2 instead of 5 for the third entry from the bottom and right. Five animals were used in each case ($n = 5$), taken at random from 350 male mice from normal stock (weights between 28 and 32 grams). Ten replicate sets (labeled as indicated at the column heads, except that primes have been added here to avoid confusion with other symbols) were injected in each case (i). Any one of the columns (A′ to K′) corresponds to an assay test. The composite test-plan has $h = 7$, $H = 10 \cdot h = 70$, $n = 5$, $n \cdot H = 350$, $R = 2$, and $d = \log 2 \doteqdot 0.30103$. The critical response was death within a definite period of observation (four days after injection).

Thus the doses form a geometric progression, and their logarithms form an arithmetic progression with a constant difference between successive values that we will represent by d. For convenient abbreviation, let

[1] $$L_i = \log D_i \text{ ; and, accordingly,}$$

[2] $$d = L_j - L_{j-1} \text{ for } j = 2, \cdots, h; \text{ whence } d = \log R \text{ and}$$
$$L_{i+u} = L_i + d \cdot u.$$

Subscript Convention. For the sake of brevity, in the preceding and other expressions where the implied meaning is obvious, let any symbol used in a subscript without an accompanying fuller definition be allowed any meaning consistent with prior definitions of other symbols in the relations involved. Thus obviously, u, above, is an integer greater than $-i$ and not greater than $h - i$. Accordingly also, it is unnecessary above to define j by the expression $j = 2, \cdots, h$; indeed, i could have been used instead of j throughout, as the equations above in which j is employed as a subscript are obviously intended to indicate relations between things already defined, not to define something denoted by L_0 nor D_0.

Common (base 10) logarithms are intended in these relations unless otherwise implied. However, it may sometimes appear convenient to use logarithms to the base R instead; as then $d = \log_R R \equiv 1$, and if one of the values of L_i is an integer in this system then all are integers in a natural succession. Thus in figure 1 the data of table 1 are represented graphically by circled points with $p_i = r_i/n_i$ as ordinate and $\log_2 D_i$ as abscissa for $i = 1, \cdots, h$ in each of ten diagrams corresponding to the columns A′ to K′ of the table.

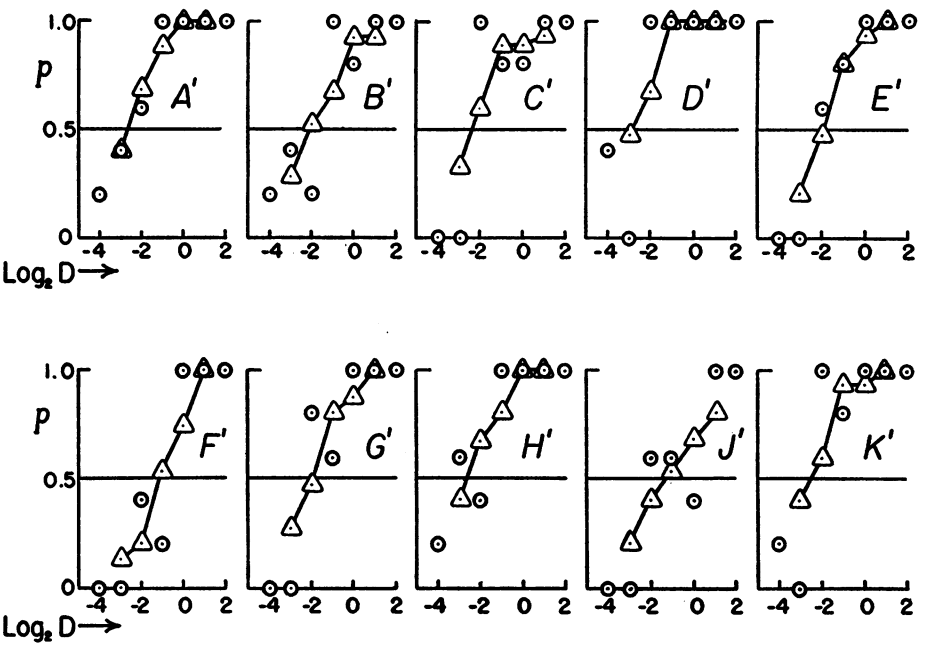

FIG. 1. Illustration of Moving-Average Estimation of Median-Effective Dose. The circles represent $p_i = r_i/n$ from the data of table 1, the triangles are corresponding moving average of three successive values of p_i, their join-lines form a polygonal curve that crosses the 0.5-horizontal at the *median-point* estimate, whose abscissal value is $\log_2 m$. The mean $\log_2 m \backsimeq -2.123$ and the standard deviation estimate $s' \backsimeq 0.588$, corresponding to a percentage deviation in m of -33 or $+50$.

Illustration of Graduating Influence. Obviously, an attempt to estimate the median-effective dose M by simple interpolation between the circled points of figure 1 would lead to erratic and occasionally equivocal results, as may be seen by joining successive circles by straight lines in each graph. If instead we take the triangle-enclosed points in these graphs, which have respective arithmetic means of three successive values of p_i as ordinate and the mean of the three corresponding values of L_i as abscissa; the join-lines as given in the graphs indicate an *ironing out* of some of the erratic variations. Where such a join-line crosses the $p = 0.5$ horizontal, also illustrated, the intersection abscissa is $\log m$ which is the estimate of $\log M$ in this case (for moving averages of *spans* of three successive values of p_i). However, calculation of $\log m$ algebraically is preferable and even simpler.

General Notions of Moving-Average Interpolation. General formulas for interpolation from moving-average points for spans of K successive values of p_i are readily developed from definitions of successive points designated by (L', p') and (L'', p''); actual graphic constructions are not required. For brevity let $b \equiv a + K$, and let

$$[3] \qquad p' = \frac{p_a + \cdots + p_{b-1}}{K}, \text{ and}$$

$$p'' = \frac{p_{a+1} + \cdots + p_b}{K};$$

and correspondingly let

$$[4] \qquad L' = \frac{L_a + \cdots + L_{b-1}}{K} = \log D', \text{ and}$$

$$L'' = \frac{L_{a+1} + \cdots + L_b}{K} = \log D''.$$

If $p' \neq p''$ and $p' \leqq 0.5 \leqq p''$, then it is possible to estimate $\log M$ (the value of $\log D$ for $\bar{p} = 0.5$, as previously defined) by a simple linear interpolation, as follows:

Let $\log m$ denote this estimate, then

$$[5] \qquad \log m = L' + d \cdot f, \text{ where } f = \frac{0.5 - p'}{p'' - p'}.$$

Calculation of $\log m$ may be simplified by use of general formulas, or simplified formulas in given situations.

General Formulas for log m. Obviously, relations [2] and [4] give $L' = L_a + d \cdot (K - 1)/2$; whence relations [3] and [5] give the general formula,

$$\log M \cong \log m = L_a + \frac{d(K-1)}{2} + d \cdot f, \text{ where}$$

$$[6] \qquad f = \left(\frac{K}{2} - p_a - \cdots - p_{b-1}\right) \Big/ (p_b - p_a) \quad \text{for } p_i = r_i/n_i,$$

$$\text{which reduces to} \quad f = \left(\frac{nK}{2} - r_a - \cdots - r_{b-1}\right) \Big/ (r_b - r_a) \quad \text{for } n_i = n.$$

Obviously, from the definition in relation [5], the result is an interpolation if and only if the fraction f lies in the unit interval $(0 \leqq f \leqq 1)$. This is equivalent in the latter form (for $n_i = n$, the usual condition of the test plan) to the requirement that $r_b \neq r_a$ and that the *r-trial* function defined as $r_a + \cdots + r_b - (nK/2)$ lie in the interval (r_a, r_b); and it is convenient to run down the conventional column of r-values, summing *comprehensive spans* of $K + 1$ successive

values and noting when this sum minus one-half nK is equal to the first or last value of r in the summation or between them. Then the index of the first value of r is a in the formula above.

Where the values of n_i are not all alike, a corresponding test (valid in general) may be made: that $p_a \neq p_b$ and the *p-trial* function, $p_a + \cdots + p_b - K/2$, lie in the interval (p_a, p_b) is a necessary and sufficient condition for the result of application of formula [6] to be an interpolation. It should be noted that the *comprehensive* (total) *index span* of data used for the estimation is $K + 1$, one more than the moving-average span K.

As mentioned previously, it is readily verified that transposition of p_i and q_i and of r_i and s_i in formula [6] gives the same value of $\log m$. The trial functions and corresponding intervals used in the test for interpolation may be altered but correspondingly so that the test yields the same conclusion. Accordingly, here as will be found elsewhere, the *sense convention* about the fundamental curve is only an aid to discussion. Furthermore, d need not be positive, and values of D_i may be replaced in the protocol by values proportional to their reciprocals (for example, by a corresponding sequence of dilutions of reagent) provided that proper interpretation of the formal result is made. However, the conventional meanings will be retained for the sake of clarity.

Simplification of Formulas in Given Situations. It seems advisable to fix K on the basis of some experience with a given system, as well as n, d, and h in the test plan, with care to reduce to a tolerable risk the chance of obtaining results that are indeterminate; although there always is some risk of this with any method of estimating median-effective dose from data of the kind considered. Once any of the parameters d, n, and K have been fixed, relation [6] may be simplified further by substitution of the given constants in the formula, and usually also by a collection of some terms.

Thus with the data of table 1, where $n = 5$ and $d = \log 2 \doteqdot 0.30103$, if we decide to use $K = 3$ as illustrated graphically in figure 1, then we may first simplify relation [6] to

$$[6.1] \qquad \log m = L_a + d\left(\frac{3-1}{2}\right) + d\left(\frac{7.5 - r_a - r_{a+1} - r_{a+2}}{r_{a+3} - r_a}\right),$$

$$[6.2] \qquad \therefore \log m \doteqdot L_a + 0.30103\left(1 + \frac{7.5 - r_a - r_{a+1} - r_{a+2}}{r_{a+3} - r_a}\right).$$

Correspondingly, the trial function is $r_a + r_{a+1} + r_{a+2} + r_{a+3} - 7.5$ under the circumstances, and its value must lie in the interval (r_a, r_{a+3}) with $r_a \neq r_{a+3}$ for the result to be an interpolation, and this is also a sufficient condition.

Example Calculations and Use of Trial Function. Consider the data of Column A′ of table 1. To try the suitability of $a = 1$ for interpolation by formula [6.2], we evaluate the trial function as $\underline{1} + 2 + 3 + \underline{5} - 7.5$ which equals 3.5, and as this lies in the interval from 1 to 5 (the first and last of the successive values of r in the summation just given) we know that an interpolation will be the result. It is readily verified that any value of a greater than unity would not satisfy the trial-function test in Column A′. Next, taking $a = 1$, we substitute the required values of r in formula [6.2] which gives, as $L_a = L_1 \doteqdot \bar{2}.7959$ from the table,

$$\log m \doteqdot \bar{2}.7959 + 0.30103\left(1 + \frac{7.5 - 1 - 2 - 3}{5 - 1}\right)$$

$$\doteqdot \bar{2}.7959 + 0.30103(11/8) \doteqdot \bar{2}.7959 + 0.4139 \doteqdot \bar{1}.210.$$

Now, consider Column E'. Successively test $a = 1, 2, \cdots$, etc. With $a = 1$ the trial function value is $\underline{0} + 0 + 3 + \underline{4} - 7.5 = -0.5$, obviously not in the interval (0,4). Trial of $a = 2$ gives instead $\underline{0} + 3 + 4 + \underline{5} - 7.5 = 4.5$, which *is* in the interval (0, 5). Trial of $a = 3$ gives $\underline{3} + 4 + 5 + \underline{5} - 7.5 = 9.5$, which is outside the interval (3, 5). Obviously, greater values of a will not meet the test condition, and use of $a = 2$ is the only procedure to yield interpolation by application of formula [6.2] to E'. Accordingly, as $L_2 = \bar{1}.0969$, we obtain the estimate

$$\log m \cong \bar{1}.0969 + 0.30103\left(1 + \frac{7.5 - 0 - 3 - 4}{5 - 0}\right)$$

$$\cong \bar{1}.0969 + 0.30103(11/10) \cong \bar{1}.0969 + 0.3311 = \bar{1}.428.$$

Estimation of Variance and Standard Deviation. In general (27, 28) for any variate X, if we have a random sample of N values (X_j), where $j = 1, \cdots, N$, we denote the sample mean by $\bar{X}$ and let v_X be the estimate of variance of X in the sampled population, where $\bar{X} = (1/N) \cdot \Sigma X_j$ and $v_X = \Sigma (X_j - \bar{X})^2/(N - 1)$. The *true* variance of X in the sampled population is denoted by σ_X^2 and the standard deviation is its square root, $\sigma_X \geqq 0$. We let $s'_X = \sqrt{v_X}$ represent the estimate of standard deviation (the notation being in accord with Rider's text (28)).

From Columns A' to K' of table 1 we obtain with $K = 3$, as indicated previously, the following ten estimates of $\log M$ successively: $\log m = \bar{1}.210, \bar{1}.360, \bar{1}.285, \bar{1}.147, \bar{1}.428, \bar{1}.669, \bar{1}.428, \bar{1}.210, \bar{1}.624$ and $\bar{1}.247$. The mean value is approximately $\bar{1}.361$, and the standard deviation of $\log m$ estimated by $s'_{\log m} \cong 0.177$.

Another method of estimating $\sigma_{\log m}$ where extensive replicate assays are not available is derived and discussed in the Appendix. The estimate may be made from data of a single assay, but it is strictly applicable only where subjects are obtained by unrestricted random sampling. Nevertheless, it furnishes an approximate indication of any gain or loss in efficiency by use of other sampling technics; for example, if stratified random sampling technic is used, it should lead to greater, if not equal efficiency (26a).

FEATURES OF MOVING-AVERAGE INTERPOLATION IN RELATION TO OTHER METHODS

Comparison of Results by Other Statistical Methods. The data of Wilson and Topley in table 1 furnish a ready basis for comparing estimations of $\log M$ by certain formulas such as relation [6] with the results Irwin and Cheeseman obtained (23, 24) from the same data by the method of Bliss (14, 15) and by that of Kärber (21, 22). By Bliss's method they estimated $\log (LD_{50})$, $\log M$ in the present notation, and obtained a mean value of $\bar{1}.361$ and standard deviation estimate $s' = 0.199$. The results by the present method with a moving-average span $K = 3$ compare favorably: mean $= \bar{1}.361$ and $s' = 0.177$. The results by these and two other methods are indicated in table 2; but, for convenience, 1000 times the difference between the estimate of $\log M$ and $\bar{1}.361$ is listed in each instance. Results by Kärber's method and by that of Reed and Muench are

thus given. With the latter an abridgement to use of a total index span of four was made, in order to correspond to use of $K = 3$ for the moving-average method. For each assay (A′ to K′) the relative order of absolute deviation from $\bar{1}.361$ is given in parentheses for the results obtained by the four rival methods, or mean values in cases of a tie. The mean of these order-numbers for each method is given also parenthetically at the base of the table.

TABLE 2

Differences times 1000 between estimates of log M and the arbitrary reference point, $\bar{1}.361$, for the ten replicate assays of table 1 by different methods. The four estimates for each assay are given order numbers in parentheses according to absolute deviation from the reference point. Results by the first two methods are those of Irwin and Cheeseman (23, 24).

ASSAY LABEL	METHOD			
	Bliss	Kärber	Reed-Muench (4-span)	Moving Average ($K = 3$)
A′	−190, (4)	−174, (3)	−164, (2)	−151, (1)
B′	−20, (3)	+7, (2)	+37, (4)	−1, (1)
C′	+30, (2)	+7, (1)	−83, (4)	−76, (3)
D′	−262, (4)	−234, (3)	−214, (1.5)	−214, (1.5)
E′	+77, (4)	+ 67, (2.5)	+37, (1)	+67, (2.5)
F′	+295, (1)	+308, (2.5)	+376, (4)	+308, (2.5)
G′	+79, (4)	+67, (2.5)	−1, (1)	+67, (2.5)
H′	−199, (4)	−174, (2)	−189, (3)	−151, (1)
J′	+311, (4)	+308, (3)	+282, (2)	+263, (1)
K′	−120, (4)	−114, (2)	−114, (2)	−114, (2)
Mean.....	0, (3.4)	+7, (2.35)	−3, (2.45)	0, (1.8)
s'	199	190	198	177

Use of $\bar{1}.361$ as arbitrary reference point in computation of the deviations might be suspected of placing the other methods in an unfavorable light, but this was the mean obtained by the moving-average and the Bliss methods. The order numbers would be the same if the grand mean estimate, $\bar{1}.362$, were used instead or if the mean, $\bar{1}.358$, from the results with the Reed-Muench method were used. If the mean from Kärber's method, $\bar{1}.368$, were taken as reference point instead, this would affect the order numbers only in an interchange of one and two in the row for B′, and the mean orders correspondingly by 0.1 as an increase for the moving-average method and decrease for that of Kärber.

Absolute deviations of estimates by the Bliss and moving-average methods for the same assay (A′ to K′) are nowhere equal and favor the Bliss method only twice in the ten trials. On the *null hypothesis* that the probability is one-half that the Bliss method would deviate less on any given trial than would the moving-average interpolation (equality excluded), that this should occur no more than twice in ten independent trials would have a probability $P = 56/1024 < 0.055$. This is almost at the ordinary critical level of significance ($P = 0.05$); moreover, the *burden of proof* in this situation lies on the opposite side. Thus the evidence of Wilson and Topley's data would not warrant use of Bliss's method instead of the *basic* (moving-average) method in the situation in question.

Comparison in Principle with Kärber's Method. Irwin and Cheeseman (23, 24), referring to the Bliss method as "the exact" and to Kärber's as "the approxi-

mate" method, used the latter in quest of a less difficult means of estimating log M; yet they conformed to precedent in use of an unnecessarily complicated computation procedure suggested by Kärber (21). As shown below, the same result may be obtained more simply. However, it appears essential for consistent use of the method to make a disagreeable assumption about the biological system beyond the actual range of observations, namely, that for any dose D_i outside the experimental range on one side $\bar{p}_i \equiv 0$ and on the other side $\bar{p}_i \equiv 1$. By the *sense convention* this would mean $\bar{p}_i \equiv 0$ for $D_i < D_1$, and $\bar{p}_i \equiv 1$ for $D_i > D_h$.

Tentatively suspend judgment of acceptability of the assumptions and disregard the issue of interpolation or extrapolation. Then, in the present notation the prerequisite assumption of Kärber's method (for $n_i = n$ as usual) may be stated as follows: there exist numbers, α and β, such that $r_i = 0$ for $i \leqq \alpha$ and $r_i = n$ for $i \geqq \beta$; furthermore i may be considered as extended indefinitely in either direction with the corresponding dosage values defined by the relation, $D_i \equiv R^i \cdot D_0$. Then, for $a \leqq \alpha$ and $b \geqq \beta$, we have $r_a = 0$ and $r_b = n$. Now, for this case in relation [6] we have

$$[7] \qquad \log m = L_a + d(K - \tfrac{1}{2}) - \frac{d}{n}(r_{a+1} + \cdots + r_{b-1}),$$

$$\text{if} \quad r_a = 0 \quad \text{and} \quad r_b = n;$$

whence, as $b \equiv a + K$, relation [2] gives

$$[8] \qquad \log m = L_b - \frac{d}{2} - \frac{d}{n}(r_{a+1} + \cdots + r_{b-1}), \quad \text{if} \quad r_a = 0 \quad \text{and} \quad r_b = n.$$

It is readily verified that this formula yields, with an indicated simpler computation, identically the same result as the method of Kärber (21–24). However, as the risks of extrapolation are disregarded, Kärber's method appears as a degenerate form of the moving-average method, even if the special assumptions are not challenged.

It is objectionable under some circumstances to extend K, the moving-average index span, more than enough to provide a stabilizing influence on the estimates of log M. This is discussed below and indicated by results in table 3. With regard to the prerequisite assumption of Kärber's method, it is obvious by use of relation [2] and simple rearrangement of terms that with any $a \leqq \alpha$ and $b \geqq \beta$ in relation [8] the same value of log m is obtained. Thus, provided this assumption is made, further extension of the index span ($K \equiv b - a$) makes no difference in the result. However, the extension may be so great in any case that only extrapolation is possible, actually so six times in ten with the data chosen for illustration by Irwin and Cheeseman (23, 24) as indicated in the last column of table 3.

A questionable common practice in actual use of Kärber's method if $r_1 \neq 0$ is to take $a = 0$ assuming $r_0 = 0$, or likewise if $r_h \neq n$ then to take $b = h + 1$ assuming $r_{h+1} = n$. Objection may be raised against an assumption that values of r_i for lower doses of toxin than used in the actual experience would all be zero in a hypothetical extension of the experience in order to provide one of the conditions necessary to application of Kärber's method. In the situation presented in table 1, this would amount to an assumption that a

temporary (four-day) immortality would have been conferred upon any and all animals injected in the prescribed way with a dose less than the least that was actually used. On the other hand, if $\bar{p}_i$ approaches a constant > 0 as i decreases indefinitely, use of Kärber's method (formally equivalent to use of relation [8]) with extension of the experience indefinitely to include actual use of lower and lower doses D_{1-j} for $j = 1, \cdots$, etc., and taking a successively lower should lead to lower and lower estimates of log M with m approaching zero as a limit. Otherwise stated, the median-effective dose of any toxin would always be estimated as no toxin at all. It seems thus that an obviously absurd result is avoided in practice only by failure to extend the actual experience to low enough doses to make the great potential influence of the sum, $r_a + r_{a-1} + \cdots + r_{a-j} + \cdots$, apparent. Such a

TABLE 3

Differences times 1000 between estimates of log M and the arbitrary reference point, $\bar{1}.361$, *for the ten replicate assays of table 1 by the method of moving averages with various index spans K; showing bias introduced with extrapolations and a counter distorting effect of substituting* $r_1 = 0$ *instead of the known values in the case of* $K = 6$, *an exaggeration of influences in Kärber's method as used by Irwin and Cheeseman (23, 24) which is equivalent here to use of* $K = 7$ *with the assumption that* $r_0 = 0$ *is applicable beyond the range of the experience.*

K⟶ / Assay Label	1	2	3	4	5	6	6 ($r_1 = 0$?)	7 ($r_0 = 0$?)
A′	−114	−114	−151	−189′	−226′	−264″	−114′	−174′
B′	+150	+87	−1	−13	−1′	−38′	+67′	+7′
C′	−114	−114	−76	−38	+7′	+7′	+7′	+7′
D′	−114	−114	−214	−314′	−415″	−515‴	−114′	−234′
E′	−13	+37	+67	+67	+67	+67′	+67′	+67
F′	+451	+388	+308	+308	+308	+308	+308	+308
G′	−76	−13	+67	+67	+67	+67′	+67′	+67
H′	−126*	−114	−151	−189′	−226′	−264″	−114′	−174′
J′	+338*	+288*	+263	+308	+308	+308	+308	+308
K′	−114	−114	−114	−114	−151′	−189″	−53′	−114′
Mean.....	+19	+22	0	−11	−26	−51	+43	+7
s'	213	183	177	206	233	261	159	190

Note: An asterisk indicates that a unique result was not obtained; the mid-range of the calculated values is given. The single (′), double (″) and triple (‴) primes after numbers indicate that the estimate was not found within the interpolation interval; but the shortest possible extrapolation was used, requiring extension respectively to the first, second or third interval beyond.

dilemma appears to confront any attempt to base a method of estimating median-effective dose upon such assumptions about the *true* probability of critical response $\bar{p}_i$ beyond the range of experiment. It appears necessary to provide some equivalent of a limitation of the range of values of D_i that are allowed to influence the estimation, objectively applied rather than as a fortuitous result of an obvious practical need for some limitations in any experiment.

Extent of Moving-Average Span (K). The results given in table 3 serve to illustrate the influence of choice of the index span of moving averages, taking $K = 1, \cdots, 6$. For convenience, as previously, 1000 (log m − $\bar{1}.361$) is listed

instead of log m. Use of $K = 1$ amounts to the same thing as simple interpolation between successive values of p_i that straddle 0.5, and whenever there is more than one such pair equivocal results are obtained. These occurred in two of the ten replicate assays (H′ and J′); they are indicated in the table and are apparent on inspection of the circled points of figure 1. Where a unique result is not obtained, the midpoint between extremes of the equivocal values is given in the table as indicated by an asterisk. With $K = 2$, the least span for actual use of moving-average graduation, only one assay (J′) gives equivocal values. With $K = 3$ the results are all unique interpolations. If we take K greater than this then estimation of log m is sometimes beyond the reach of interpolation from the available data (table 1), but the least possible extrapolation is used. This is indicated in table 3 roughly by single, double or triple primes to signify respectively that extrapolation was required involving a reach to the first, second, or third equal interval (of length d) beyond that corresponding to interpolation. With $K = 4$ such effects are of a relatively minor nature, and are encountered in only three of the ten assays (A′, D′, and H′). As K is increased beyond four, extrapolation is required more frequently and appears to be more influential in introducing bias into the mean estimates of log M; a trend toward lower mean values of log m is noticeable. Precision, as indicated by the standard deviation estimates s', also seems best with $K = 3$.

Extended K and Assumptions of Kärber's Method. If the assumption were correct, as made by Irwin and Cheeseman in applying Kärber's method to the data of table 1, that smaller doses such as $D_0 = 0.03125$ mg of the toxin preparation would have yielded no deaths in response and thus $r_0 = 0$ in every assay (A′ to K′); then we might use $K = 7$ as we have already used $K = 1, \cdots, 6$. The results thus formally obtained by use of the reduced form of relation [6] given in relation [8] for the special case ($r_a = 0$, and $r_b = n$) are necessarily the same as those obtained by Kärber's method, but are given again (table 3) to emphasize a failure to find in this last member ($K = 7$) evidence of family traits apparent in the six preceding offspring. A sudden reversal of the trend in mean log m is a striking result: having successively, with $K = 1, \cdots, 6$, obtained mean values of 1000 (log m − $\bar{1}.361$) = +19, +22, 0, −11, −26, −51; we obtain the value, +7, for $K = 7$ with the questionable assumption that we should have found $r_0 = 0$ throughout if D_0 had been used. There is a corresponding reversal of trends in frequency and extent of extrapolation and in the estimate s' of standard deviation, all ostensibly favorable but damningly contrary to evident family traits. In the next to last column of table 3 are presented results of an exaggeration of the suggested influence, obtained with $K = 6$ by substituting zero for the actual values of r_1 which is equivalent to assumption that $r_1 = 0$ throughout in ignorance of the actual data (obviously false five times in ten).

The evidence suggests that apparently good results obtained by Kärber's method may be due to compensating errors of the sort introduced in the present case by a subjective judgment that r_0 should be zero and by use of an excessive index span ($K = 7$).

Characteristics of Indefinite Cumulant and Reed-Muench Methods. Objections, closely resembling those made to Kärber's method, appear to discourage any attempt to base methods upon two indefinite cumulants, the sum of all values of r_i for $i \leqq c$ and the sum of all values of s_i for $i \geqq c$.

Briefly these sums may be denoted by $\sum_{c} r$ and $\sum_{c} s$ where

$$[9] \qquad \sum_{c} r = \sum_{i=1}^{c} r_i \quad \text{and} \quad \sum_{c} s = \sum_{i=c}^{h} s_i$$

provided it is assumed that $r_i = 0$ for $i < 1$ and that $s_i = 0$ for $i > h$. The methods employ

$$[10] \quad \phi_c' = \sum_{c} r / (\sum_{c} r + \sum_{c} s) \text{ and } \phi_c'' = \sum_{c'} r / (\sum_{c'} r + \sum_{c'} s) \text{ where } c' = c + 1$$

to obtain an estimate m' of the median-effective dose M by linear interpolation between points, $(\log D_c, \phi_c)$ and $(\log D_{c'}, \phi_c'')$, to find the "endpoint", $(\log m', 0.5)$. It is noteworthy that here $\phi_c'' = \phi_{c'}'$. Obviously, the results of this procedure could be greatly biased by not having $n_i = n$, a constant; although substitution of p_i for r_i and q_i for s_i throughout would remove this objection.

Reed and Muench (25) adopted the condition $n_i = n$, and specified that the data be abridged so that the calculation of the "endpoint" is based on data from an equal number of dosage values (D_i) on each side of it. This specification is essentially equivalent to limiting the range of the index i in the summations to the inclusive interval $(c + 1 - k, c + k)$ for an arbitrary integer k, provided that c is such that $\phi_c' \leqq 0.5 \leqq \phi_c''$. Strictly interpreted, the equality signs in the last expression admit an exception to the provision of Reed and Muench, ". . . an equal number of dilutions is taken on each side of the endpoint." However, their statement (25) is immediately followed by an example that appears to condone even further departure from a strict interpretation of the provision, although the data there given would have yielded the same result either way.

The Reed-Muench method appears to be placed in the most favorable light on the suggested basis, which may be given explicitly as follows:

$$[11] \qquad \phi_c' = \frac{r_{c+1-k} + \cdots + r_c}{r_{c+1-k} + \cdots + r_c + s_c + \cdots + s_{c+k}}, \qquad \text{and}$$

$$\phi_c'' = \frac{r_{c+1-k} + \cdots + r_{c+1}}{r_{c+1-k} + \cdots + r_{c+1} + s_{c+1} + \cdots + s_{c+k}};$$

and the median-effective dose M is estimated by the "endpoint" value m', given by

$$[12] \qquad \log m' = \log D_c + d \cdot \frac{0.5 - \phi_c'}{\phi_c'' - \phi_c'}, \quad \text{provided} \quad \phi_c' \leqq 0.5 \leqq \phi_c''.$$

Thus, if the range of the index i is great enough in the protocol, for a given value of k there may be determined a succession of intervals, (ϕ_c', ϕ_c''), one of which may be found to contain 0.5 and thus serve in evaluation of m' in relation [12]. Unfortunately, the successive intervals, (ϕ_c', ϕ_c'') and $(\phi_{c'}', \phi_{c'}'')$ where $c' = c + 1$, overlap except in the case where $r_{c+1-k} = s_{c+1+k} = 0$ and in two *trivial* other cases where ϕ_c'' and $\phi_{c'}'$ are both zero or both one. This is readily verified from the definitions in [11]. Thus there is an almost universal tendency for the Reed-Muench method to give biased or equivocal results, the latter occurring whenever more than one of the successive intervals contains 0.5.

A simple example is afforded by consideration of a uniform trend, the absolute antithesis of erratic data. Thus suppose that $r_i = i$ for $i = 1, \cdots, n - 1$; and $s_i = n - r_i = n - i$. In this hypothetical assay suppose that the doses are successively doubled ($d = \log 2$).

Let M' be the geometric mean of the doses (D_i) for $i = 1, \cdots, (n-1)$; then apparently M' is the best estimate of M that can be made in this situation on any basis. This is the value that would be given by the moving-average method with any value of the index span K and interpolation (or even extrapolation from any admissible part of the data). The same result would be obtained from all the data by the methods of Bliss, Berkson, and Kärber (provided that in the last method we use $r_0 = 0$ and $r_n = n$, in extension of the data). However, the Reed-Muench method gives equivocal estimates (m'). The simplest way to present the results is to give the ratio $g' = m'/M'$. Thus for a total index span $2k = 4$ we have: if $n = 10$, then equivocally $g' \cong 0.823$ or 1.215; and if $n = 15$ then $g' \cong 0.617, 1.000$, or 1.621. Similarly, for a total index span $2k = 6$ we have: if $n = 10$, then $g' \cong 0.878$ or 1.138; and if $n = 15$, then $g' \cong 0.683$, 1.000, or 1.463. The purpose here, of course, is to illustrate the character of the equivocal results, not to indicate the magnitude of relative differences to be expected in actual use.

In application of the Reed-Muench method to any actual data, following the procedure indicated by the authors, more than one value for m' would not ordinarily be obtained, because the first one found would be accepted. A limitation placed experimentally upon the range of values of D_i might prevent recognition of the danger, just as might ignoring part of a more extensive set of results (deleting data for successive dosage values at either the top or bottom of the conventional protocol). Obviously, this would be done in either case at the expense of introduction of some bias according to what range of doses is allowed to influence estimation of the median-effective dose M, the estimate m' tending to be too near the geometric mean of the range of doses so employed. Replicate assays made with the same range of doses would tend to develop in the observer a false confidence in estimates as a result of apparent precision, if he were not aware of this artificial centrifugal tendency in m'. Hedén (49), in a recent article suggesting application of the Reed-Muench method to serologic titrimetry, has given a table in which he cautions against some of the most prominent points of danger.

Road to Further Modifications of Cumulant Method. It is evident that the Reed-Muench modification of the cumulant method evokes new defects in place of those it removes. The new sources of difficulty are a failure in definition to make ϕ_c'' identical with ϕ_{c+1}', and to have ϕ_c' such that the range of the index values involved extends equally far to each side of c, if D_c is taken as the associated dosage value used in the interpolation. D_c would then be the geometric mean of dosage values corresponding to values of r_i and s_i involved in the definition of ϕ_c'. This would avoid the vicious overlapping and any obvious tendency toward biased results. However, it seems preferable to define the cumulants and ϕ_c' in terms of p_i and q_i rather than r_i and s_i, for the sake of greater generality; the other forms are readily obtained where $n_i = n$ throughout by substitution of the identical values, $p_i = r_i/n_i$ and $q_i = s_i/n_i$.

Thus, we might redefine the bases of interpolation, ϕ_c' and ϕ_c'', by

$$[13] \qquad \phi_c' = \frac{p_{c+1-k} + \cdots + p_c}{p_{c+1-k} + \cdots p_c + q_c + \cdots + q_{c-1+k}}, \quad \text{and} \quad \phi_c'' = \phi_{c+1}'.$$

The "endpoint" value m' could be defined by relation [12] in these new terms.

However, this is not the only possible recourse. Indeed, it is not essential to a use of cumulants that ϕ_c' be a cumulant ratio as in [10], [11], or [13]; but we should agree to the condition that henceforth $\phi_c'' = \phi_{c+1}'$ *identically*. In the last case [13] the underlying idea is, in accord with that of other cumulant methods, to find or estimate by some form of interpolation a dose that should be expected to yield a zero cumulant difference; i.e., the

p-cumulant minus the q-cumulant should be zero. This suggests redefinition of ϕ_c' as some convenient linear function of the cumulant difference, $p_{c+1-k} + \cdots + p_c - (q_c + \cdots + q_{c-1+k})$. Furthermore, as $q_i = 1 - p_i$ identically, we note that

$$[14] \quad p_{c+1-k} + \cdots + p_c - (q_c + \cdots + q_{c-1+k}) = p_{c+1-k} + \cdots + p_{c-1} + 2p_c + p_{c+1} + \cdots + p_{c-1+k} - k.$$

Accordingly, it would appear convenient to *redefine*

$$[15] \quad \phi_c' = \frac{p_{c+1-k} + \cdots + p_{c-1} + 2p_c + p_{c+1} + \cdots + p_{c-1+k}}{2k}.$$

As agreed, ϕ_c is defined as identical with ϕ_{c+1}'. Interpolation formally by relation [12] gives the same "endpoint" as that for a cumulant difference of zero as suggested above.

Formula [15] may be recognized as that for a weighted moving average, many different forms of which have been employed as graduation formulas (29a). It differs from the simple moving average designated by p' in relation [3] only in giving double weight to p_c in the middle term of the numerator, and in a limitation to use of odd index spans, $K = 2k - 1$. The denominator, of course, is the sum of the weight coefficients. If we define

$$[16] \quad \theta = \frac{(k - p_{c+1-k} + \cdots + p_{c-1} + 2p_c + p_{c+1} + \cdots + p_{c-1+k})}{p_{c+k} + p_{c+1} - p_c - p_{c+1-k}},$$

then the interpolation (provided $0 \leqq \theta \leqq 1$) is given by

$$[17] \quad \log m'' = \log D_c + d \cdot \theta,$$

where m'' is the estimate of the median-effective dose M. It should be noted that the test for interpolation and the calculation are more difficult than in use of the simple moving-average formula [6]. Furthermore, the restriction to use of an odd index span ($K = 2k - 1$) is awkward; and no compensating advantage is apparent.

An Equivalence of Simple Moving-Average and Modified Cumulant Methods. If we take a further modified cumulant difference, $p_{c+1-k} + \cdots + p_c - (q_{c+1} + \cdots + q_{c+k})$, and associate this with the geometric mean of dosage values involved in obtaining these data (which is equivalent to what was done in the preceding modifications); then this dosage value is readily found to be $\sqrt{D_c \cdot D_{c+1}}$. Then, proceeding in steps analogous to those leading from relation [14] to [16] and [17], we are led to relations identical with formula [6] for the simple moving-average method except that $K = 2k$ and thus is always *even*. Now, if we take another modification of cumulant difference, $p_{c+1-k} + \cdots + p_{c-1} + \frac{p_c}{2} - \frac{q_c}{2} - (q_{c+1} + \cdots + q_{c-1+k})$, and associate this with the geometric mean dose involved (D_c); again we are led by analogous steps to the same formula [6], but this time find that the index span is always *odd*, $K = 2k - 1$. Thus it appears that a process of successive modification of cumulant methods, in attempts to remove evident defects and avoid unjustified awkwardness, leads directly to the simple moving-average method that has been proposed above on the basis of immediate intuitive appeal.

The approach from the point of view of cumulant differences lies along an easily followed trail as far as the next to last modification. There the appearance of slight difficulty in the associated dosage value ($\sqrt{D_c \cdot D_{c+1}}$) might have discouraged further progress as might

the apparent introduction of complications with the companion modified cumulant difference; but, only a little further along, all such difficulties are left behind as we arrive at the same situation previously attained by the simpler approach.

Use of Elaborate Graduation Formulas and Curve Fitting. In the special case where $k = 2$ in relations [16] and [17] we have the equivalent of application of simple moving averages for $K = 2$ to like moving averages in [3], i.e., a double graduation. This and many other forms of weighted moving average as well as other graduation formulas (29a) might be used advantageously in some situations. Indeed, methods that involve fitting of curves of given type (3–20, 30) may be regarded as elaborations of such methods, useful to replace a basic method such as that now proposed, where there is good reason to expect use of the more elaborate method to lead (without introduction of intolerable bias) to important improvement in precision or to permissible abbreviation of tests (3–12). Of course, other percentiles may be used as "endpoint" instead of the median, but usually less efficiently. However, it may not be readily apparent what type of curves should be useful. Wilson and Worcester have thus been led to consider (30) a generalization of the curve fitting problem in bioassay.

The investigations by Hussey and associates (31–35) of effects of radiations on biological systems furnish an illustration (5a), both for quantal and gradational assay, of strange characteristics that may be encountered in a dosage-response curve. Furthermore, these studies provide an interesting example of a mistaken attempt (by another writer) to approximate the curve involved by fitting a fairly simple increasing function of the irradiation interval (t) to the average duration of the prepupal stage (ϕ), employing data of the first report. This is discussed in one of the later papers (33), further exploration having clearly demonstrated that such dosage-response curves on the contrary rise to a maximum value of ϕ and then fall to an almost level plateau in all cases investigated (32, 33). In the later article (33) the median is used as average in preference (5a, 13, 36) to the mean, and in the second text-figure (reproduced elsewhere (5a)) one experience is illustrated by a combined graph of median response ϕ and a class-frequency diagram of individual responses corresponding to each irradiation period t that was employed. A *quantal* assay system could be imagined as based on responses found up to any convenient value of ϕ (the abscissa of the graph); but it is easy to see how an unfortunate choice, say $\phi = 7.5$ days, would lead to ambiguities in tests and insensitivity with change in dosage above 160 minutes under the given conditions.

That difficulties of this sort may be encountered in immunologic assay is illustrated in the studies of mouse protection tests by Goodner and Horsfall (37). With certain pneumococcus antisera given to test animals in varying amounts to counter a standard dose of pneumococcus culture (sufficient to kill practically all animals otherwise) they found that over part of the range increased protection appeared to result from increased antiserum dosage, but with a further increased dosage a maximum protective effect apparently was passed, for beyond a certain point increased dosage appeared to give less protection. If such a system must be used in assay, then care must be taken to choose conditions so as to deal with the required branch of the fundamental curve or change conditions so as to eliminate the peculiarity.

An interesting case in point is presented by some of the complement-fixation systems. We have indicated that the amount of complement M required for 50 per cent hemolysis is a linear function of the amount of a given antiserum used in a reaction with the essentially *optimal amount* of antigen. With some systems (e.g., the tuberculosis tests) it is enough that the amount of antigen be in considerable excess, but with others (e.g., tests for syphilis or pneumococcus antibody) there is an amount of antigen that gives a maximal fixation

reaction with the given amount of antiserum and less fixation is obtained with either more or less antigen. Fortunately, in the syphilitic system the curve does not have a sharply defined maximum but a fairly broad range in which nearly maximal effects are obtained. Still more fortunately, the maximally effective amount of antigen depends essentially on the required maximal M, which in turn is approximately dependent on the amount of antibody present, and not capriciously on the particular qualities of the serum otherwise. Accordingly we may set up (as in the routine test for syphilis) three tubes containing respectively 3, 6, and 12 units of complement and in each the amounts of antigen appropriate if the value of M falls within range of estimation from the resulting p found.

Features of Weighting Systems in Curve Fitting. At this point it seems appropriate to examine at least the approximate influence of weighting systems commonly employed in the curve-fitting process, but in approaching this subject, perhaps like Berkson (19), we should pray for guidance. At least we may hope that the discussion will serve as a stimulant to others. As previously, we take log D as abscissa in both the fundamental and transformed coordinate systems. Accordingly, let w be the weight assigned to the squared vertical deviation of a point (observed or "corrected") from the fitted line, and assume we are to minimize the sum of the weighted squares in the fitting process.

The $T_{(p)}$ transformation is equivalent to a one-way stretching of the coordinate plane vertically in either direction from the $p = 0.5$ horizontal. The factor for the total *stretching* from the transformation of a point not on that horizontal is $[T_{(p)} - T_{(0.5)}]/(p - 0.5)$. As previously, it is convenient but not necessary to have $T_{(p)}$ an increasing function of p; then this total-stretch factor is positive. However, the local distortion or *stretching factor*, which we shall denote by $1/E$, is given by the derivative of $T_{(p)}$ with respect to p. Thus, with base e logarithms if $T_{(p)} = \log \frac{p}{1-p}$, then $E = pq$; or if $T_{(p)} =$ either the normal deviate or the probit of p, then E is the corresponding ordinate (z) of the *normal* curve. Use of logarithms to any other base (say a) instead with $T_{(p)} = \log \frac{p}{1-p}$ would give $E = pq \cdot \log_a e$, a constant times the value $E = pq$ for natural logarithms. Of course, any set of weighting coefficients may be multiplied by any constant throughout without essential alteration in the result of the curve fitting.

Now, let w' represent a weight we might agree to give to squared deviations in p from the fundamental (log D, p) curve approximation, e.g., w' might be taken as the reciprocal of the estimated variance of p in random sampling, then $w' = n/(pq)$. Hence we might expect $w = E^2w'$, and approximately this is the weighting system used by Bliss (14) and Berkson (19, 20) with the appropriate interpretation of E, although adjustment of the transformed data may be involved also.

Obviously, the corresponding negative transformation, $-T_{(p)}$, gives the same weighting coefficient $w = E^2w'$. Thus for the logit transformation the weighting coefficients may be taken as $w \cong np^2q^2/(pq) = npq$. It is true here and also, but less markedly, with the probit system that less weight is given to points in the *transformed* coordinates the further p is from 0.5; but we should guard against the fallacy of supposing that therefore original points (log D, p) are given less

weight according as p is remote from 0.5. Indeed, the opposite is clearly evident from the derivation as Garwood (16) and Berkson (20) have emphasized.

As has been noted, the difficulty with $p = 0$ or $p = 1$ in the observations has been treated (14, 15, 20) by adjustment of data. Strictly in accord with the principle applied, all points should be adjusted (14). The adjustment is based on a straight line fitted by inspection or otherwise to the transformed points or successive approximations by straight lines generated from that beginning. The "corrected" or "fictitious" points employed by Bliss (14) and Fisher (15) are taken so that the weighted sum of squared deviations from the provisional line is $\sum [n_i(p_i - p_i')^2/(p_i'q_i')]$, where p_i' is the corresponding p-value on the provisional line and $q_i' = 1 - p_i'$; and the process is aimed at finding a line which makes this sum a minimum, and is essentially equivalent to fitting an integrated normal curve (16) to the points in the fundamental form so as to minimize this sum. The logistic function instead of the integrated normal curve may be employed in a similar manner (17, 18, 20).

It is at least of theoretical importance to note that, apparently, the conditions for convergence of the curve-fitting process have not been carefully specified. Good examples have been given, especially by Garwood (16), to suggest the nature of the processes involved; but a general theorem of convergence has not been proved for any set of points $(\log D_i, p_i)$, or with certain specified exceptions, where we are required to fit a given curve or a straight line in transformed coordinates in the indicated manner. Indeed, the general theorem without such exceptions can easily be disproved, and for this it is sufficient to cite a single example of failure. Thus for any set of doses, $D_1 < D_2 < D_3$ or more specifically $D_3 = 2D_2 = 4D_1$, let the values of p_i be given by $p_1 = 0 < p_2 < 1$, and $p_3 = 1$. In this example the method of Bliss and Fisher would not direct us toward any line of finite slope and the estimated median-effective dose would appear to be D_2. A second example with just the first two points instead would yield essentially the same result, the estimate of M being D_2. In either case, this would be so regardless of the value of p_2.

The present purpose is not to determine the limitations of these methods, but to indicate that the issue should be investigated at least to the extent of specification of some definite *practical limitations*. Of course, no method of estimation of median-effective dose should be applied to data where p_i is constant in the given experience (49), nor if for a given i we have $p_{i+j} = p_{i-j}$ throughout the data utilized, nor without restriction of range to exclude the influence of a possible indefinite extension of dosage values with the *true* $\bar{p}_i$ approaching asymptotically some value other than zero or one. However, the examples cited above are not of that type and demonstrate that other exceptions are required. An immediately suggested expedient might be based on the fairly common practice of excluding all values of p_i that are zero or one in obtaining the first provisional line and requiring thus that at least two other values of p_i remain. Then we might proceed to obtain the second provisional line with exclusion of all points for which the *expected value* p_i' (taken from the preceding line) is not in the interval (0.001, 0.999), and to obtain any successive fitted lines with exclusion of points for which p_i' on the immediately preceding line is outside (0.05, 0.95) or some other preassigned interval. Obviously, at least two points must remain to yield the line. Such a system might overcome some of the indicated difficulties, and appears to offer only an objective direction of what might be expected to result from good judgment.

Another difficulty, not apparent in the mathematics based upon the original assumptions, arises as a result of infringement of one of the assumptions, made at least tacitly, that other sources of variation in the observations (log D, p) are negligible with respect to the variation in p from random sampling. Perhaps we might stretch the condition to

include cases where the other sources of variation are no more influential (48). However, as Neyman (12c) has pointed out, this is not necessarily so. Indeed, in any practical situation, this assumption implies that n is not so great as to make the variance of p negligible with respect to contributing variance arising from other sources; and, as contrary cases, we have previously cited the complement-fixation systems where $n \geqq 10^8$ and thus $2\sqrt{pq/n} \leqq 0.0001$. Such conditions effectively undermine the basis on which there has been established a preference for taking deviations in the given direction. This is not enough to establish the contrary preference, but certainly throws the question open again, and in such cases invalidates the weighting system described above.

Still another difficulty lies in the use of the *index of dispersion*, $\Sigma[n_i(p_i - p'_i)^2/(p'_i q'_i)]$, as the basis for approximating a *maximum likelihood* fit. It is well known (27, 28) that, especially for small values of n_i and for p'_i remote from 0.5, the approximations involved may be poor. Accordingly, with any given data any method so based, such as that of Bliss and Fisher (14–16), should not be considered even to *aim* at an *exact* maximum likelihood solution, and reference to it as "the exact" method (23) even in quotation marks may lead to a false appreciation. However, this deflection of aim does not appear to detract from the practical excellence of these curve-fitting methods when suitably applied.

Moving-Average Interpolation Viewed as a Basic Method. In certain situations where it may seem desirable to compare the efficiency of a given method of estimating the median-effective dose M with that of a *basic* method of simple character, the simple moving-average interpolation may well serve the purpose. It possesses the most desirable basic characteristics: independence of assumption about the precise form of the *fundamental* (logarithm of dosage, relative response frequency) curve, and use of well-known principles of graduation and interpolation in the estimation. As the necessary calculations are simple, it may be regarded as the method of choice unless another is shown considerably more efficient in a given situation. A choice has to be made here, as elsewhere (36a), between methods which seem to rest on a relatively secure base and others which involve an apparent risk in further assumption but offer a prospect of gain in either power or economy or both. Comparison of a more aggressive method with a method more *basic* in the given sense may seem entirely unnecessary if other evidence of value may be obtainable (11, 12, 44, 45) as has been mentioned in the discussion of complement-fixation tests. Caution is desired but not timidity.

Of course, the fundamental curve should be approximately symmetrical with respect to the *median-point* (log M, 0.5) at least within the interval (log D_a, log D_b) to be well-suited to estimation of that point by the linear interpolation from simple moving averages. However, by suitable limitation of range (made possible by choice of d, n, and dosage range in the test plan and use of an appropriate value of K) good conditions for estimation of M in this way should be attainable in almost any situation where the notion of median-effective dose is useful. Indeed, its definition is based essentially on the supposed possibility of its estimation by the method of simple interpolation, equivalent to the special case $K = 1$ in formal use of relation [6]. The types of curves (logistic and integrated normal) most used as approximations of the *fundamental curve form* are S-shaped symmetrically about the estimated median-point (log M', 0.5); thus even greater emphasis is placed on a condition that is favorable to use of

the moving-average method. Furthermore, as Berkson (20) has pointed out and we have indicated above, both his own method and that of Bliss and Fisher operate approximately as if to fit a curve (respectively, the logistic and the integrated normal) to the data in the fundamental form of points (log D_i, p_i) with a *weight* proportional to $n_i/p'_iq'_i$ assigned to squared deviations, $(p_i - p'_i)^2$, where p'_i is the ordinate of the point (log D_i, p'_i) *on the fitted curve.* Thus, these methods give relatively more weight to the original (log D, p) points according as they are more remote from the estimated median-point (log M', 0.5). The simple moving-average interpolation method does not do this. Equal weight is given to all points utilized in the actual calculation except the first and last, which are given weights no greater—usually somewhat less—automatically dependent on the value of f as may be seen in the Appendix; and all other points have no weight except in the preliminary choice of the range (a, b) to be used for the index i in the interpolation. This feature of the proposed basic method seems at worst to err on the side of safety. It should tend to make comparative tests sensitive to any increase in efficiency resulting from use of a curve-fitting method that places special emphasis upon the data of points remote from the median-point.

Finally, the writer wishes to make it very clear that he has no intention of discouraging use of the curve-fitting methods, which appear to have found many valuable applications.

APPENDIX

ESTIMATES OF ASSAY VARIANCE, BASED ON INTRACLASS DATA

We have noted under relation [6] that as long as f lies in the *unit interval* the same value of a is suitable for the required interpolation; i.e., an increment (Δf) in f produces d times as great an increment in log m, $\Delta \log m = d \cdot \Delta f$, where d is a constant as previously defined.

Furthermore, the partial derivative of f with respect to p_i is readily obtained for any i. From relation [6], for $p_i = r_i/n_i$, we have

$$[A1] \qquad f = \left(\frac{K}{2} - p_a - \cdots - p_{b-1}\right) \Big/ (p_b - p_a), \quad \text{and}$$

$$f - 1 = \left(\frac{K}{2} - p_{a+1} - \cdots - p_b\right) \Big/ (p_b - p_a).$$

Therefore, the partial derivatives are:

$$\frac{\partial f}{\partial p_i} = \frac{-1}{p_b - p_a} \qquad \text{for } a < i < b; \text{ and}$$

$$[A2] \qquad \frac{\partial f}{\partial p_a} = \frac{\partial (f-1)}{\partial p_a} = \frac{\frac{K}{2} - p_{a+1} - \cdots - p_b}{(p_b - p_a)^2} = \frac{f-1}{p_b - p_a}, \text{ and}$$

$$\frac{\partial f}{\partial p_b} = \frac{\frac{K}{2} - p_a - \cdots - p_{b-1}}{-(p_b - p_a)^2} = \frac{-f}{p_b - p_a}.$$

Let $\Delta p_i = p_i - \bar{p}_i$; then Δf may be approximated by

$$[A3] \quad \Delta f \cong \frac{-1}{\bar{p}_b - \bar{p}_a} \cdot [(1 - \bar{f}) \cdot \Delta p_a + \Delta p_{a+1} + \cdots + \Delta p_{b-1} + \bar{f} \cdot \Delta p_b],$$

where $\bar{f}$ is the value of f for the case where p_i is replaced by $\bar{p}_i$ throughout formula [A1]. Obviously, the *expected value* of Δf is approximately zero, $\overline{\Delta f} \cong 0$; whence the expected value of $(\Delta f)^2$ is approximately the variance of f, $\overline{(\Delta f)^2} \cong \sigma_f^2$. Assume that Δp_i is independent of Δp_j for $i \neq j$. Then the expected value of the cross-product is given by

$$[A4] \quad \overline{(\Delta p_i)(\Delta p_j)} = 0 \text{ for } i \neq j$$
$$= \sigma_{p_i}^2 \text{ for } i = j.$$

Accordingly, the variance of f may be estimated by

$$[A5] \quad \sigma_f^2 \cong \frac{1}{(\bar{p}_b - \bar{p}_a)^2} \cdot [(1 - \bar{f})^2 \cdot \sigma_{p_a}^2 + \sigma_{p_{a+1}}^2 + \cdots + \sigma_{p_{b-1}}^2 + \bar{f}^2 \cdot \sigma_{p_b}^2].$$

By definition, $s_i \equiv n_i - r_i$ and $q_i \equiv s_i/n_i \equiv 1 - p_i$; and, accordingly, $\bar{q}_i \equiv 1 - \bar{p}_i$. The *true* variance of p_i is $\sigma_{p_i}^2 = \bar{p}_i \cdot \bar{q}_i/n_i$ as is well known; and this variance may be estimated from the sample value (p_i), for $n_i > 1$, by

$$[A6] \quad v_i \equiv p_i \cdot q_i/(n_i - 1).$$

This is also well known in a more general form; but is readily verified in this special case (temporarily dropping the index i) from relations based on the point binomial, as follows:

$$n^2(n - 1) \cdot \bar{v} = \sum_{\alpha=0}^{n} \binom{n}{\alpha} \cdot \bar{p}^\alpha \cdot \bar{q}^\beta \cdot \alpha\beta$$

[A7] where $\beta = n - \alpha$; whence, for $n > 1$,

$$n \cdot \bar{v} = \sum_{\alpha=1}^{n-1} \binom{n-2}{\alpha-1} \cdot \bar{p}^{\alpha-1} \cdot \bar{q}^{\beta-1} \cdot (\bar{p}\bar{q}) = \bar{p} \cdot \bar{q}.$$

Accordingly, the expected value of v_i is $\bar{v}_i = \bar{p}_i \cdot \bar{q}_i/n_i = \sigma_{p_i}^2$, the *true* variance of p_i . Thus, as the variance of $\log m$ is approximately d^2 times that of f, we obtain from [A5] and [A6] the estimation formula,

$$[A8] \quad \sigma_{\log m}^2 \cong \frac{d^2}{(\bar{p}_b - \bar{p}_a)^2} \cdot [(1 - \bar{f})^2 \cdot v_a + v_{a+1} + \cdots + v_{b-1} + \bar{f}^2 \cdot v_b];$$

wherein, unfortunately, there still remain three quantities ($\bar{p}_b$, $\bar{p}_a$, and $\bar{f}$) that are not given by the assay experience, except that they may be approximated by the corresponding sample values (p_b , p_a , and f). However, the additional error in approximation thus induced may be so small as to be negligible, if $\bar{p}_b \cong 1$ and $\bar{p}_a \cong 0$, in comparison with other sources of error in the approximation [A8]. Estimates of variance obtained in this way are not adapted to such refined tests of significance (27, 28) as is the variance estimate s'^2 defined in the text in accord with Rider's (28) notation. Of course, if replicate assays have been made under essentially the same conditions, then composite estimates may be made or data may be pooled under some circumstances for the purposes.

In the case of constant $n_i = n > 1$, relation [A8] may be reduced to

$$[A9] \quad \sigma_{\log m} \cong \frac{d}{r_b - r_a} \cdot \sqrt{\frac{(1 - f)^2 \cdot r_a s_a + r_{a+1} s_{a+1} + \cdots + r_{b-1} s_{b-1} + f^2 \cdot r_b s_b}{n - 1}}$$

where, of course, the standard deviation estimate is positive or zero; the same value is given in this case ($n_i = n$) with r_i replaced by p_i and s_i by q_i throughout.

However, in a consideration of possible test-plan improvements, with regard to the influence of choice of values for n, d, and the moving-average index span ($K = b - a$), it is convenient to have in mind the form derived by substitution of the hypothetical true variance, $\bar{p}_i\bar{q}_i/n$, for $\sigma^2_{p_i}$ in relation [A5]:

$$[A10]\ \sigma_{\log m} \cong \frac{d}{\bar{p}_b - \bar{p}_a} \cdot \sqrt{\frac{(1 - \bar{f})^2 \cdot \bar{p}_a\bar{q}_a + \bar{p}_{a+1}\bar{q}_{a+1} + \cdots + \bar{p}_{b-1}\bar{q}_{b-1} + \bar{f}^2 \cdot \bar{p}_b\bar{q}_b}{n}}.$$

All the indicated approximations are asymptotic as $n_i \to \infty$, for $i = a, \cdots, b$. As indicated in the main text, these formulas are applicable only where unrestricted random sampling is employed or where an estimate of relative efficiency in use of other sampling technics is to be made in default of control experience with unrestricted random sampling. In the latter case direct estimates (s'^2) of variance of log m from replicate assays with the given sampling technic are required for the comparison; they are desirable in any case. The indirect estimates obviously are insensitive to technical errors such as those arising in use of a syringe, or in failure to have comparable environmental conditions at the time of administration of the test dose to subjects and thereafter to the end of the observation interval. Many such influences are likely to be greater when assays are not performed simultaneously. Nevertheless, non-simultaneous replicate assays may be preferred in order to avoid too great reliance on results that may be abnormally affected by a temporary condition of the colony of test subjects. An example might be given by replicate assays where two reagent solutions are tested for relative potency, ostensibly of a given agent. There may be other materials present in different relative amount in the two solutions and sensitivity of test animals (subjects) may differ from time to time, not only with respect to the agent in question but with respect to the other materials also. This would be an unfortunate situation, of course, but to remain ignorant of it would be more unfortunate. Accordingly, non-simultaneous replicates may seem preferable if such influences are suspected. Likewise, it seems usually preferable to compare reagent preparations with a standard (4, 5, 11, 12) agent rather than accept the reactions of a colony of subjects as a standard.

REFERENCES

1. WINSOR, C. P. 1932. A comparison of certain symmetrical growth curves. J. Wash. Acad. Sci., **22**, 73–84.
2. VON KROGH, M. 1916. Colloidal chemistry and immunology. J. Infectious Diseases, **19**, 452–477.
3. WADSWORTH, A., MALTANER, E., AND MALTANER, F. 1931. The quantitative determination of the fixation of complement by immune serum and antigen. J. Immunol., **21**, 313–340.
4. WADSWORTH, A. B. 1939. Standard Methods of the Division of Laboratories and Research of the New York State Department of Health; 2d ed. Baltimore, Williams & Wilkins. 681 p.
5. WADSWORTH, A. B. 1947. Standard Methods of the Division of Laboratories and Research of the New York State Department of Health; 3d ed. Baltimore, Williams & Wilkins, (a), p. 145–161.
6. MALTANER, F. AND MALTANER, E. 1940. The quantitative determination of the antigen-antibody reaction by complement fixation. Third Inter. Congr. Microbiology, N. Y. Rept. proc., p. 781–782.
7. THOMPSON, W. R., AND MALTANER, F. 1940. On the construction of graphs and tables for evaluation of the quantitative complement-fixation reactions and reaction ratios. J. Immunol., **38**, 147–157.
8. RICE, C. E. 1942. Studies of antipneumococcal serum. 2. Complement-fixing activity of antipneumococcal rabbit-serum with homologous type-specific carbohydrate.

Technic of test. General quantitative relationships among reagents. J. Immunol., **43**, 129–148.

9. KENT, J. F., BUKANTZ, S. C., AND REIN, C. R. 1946. Studies in complement fixation. 1. Spectrophotometric titration of complement; construction of graphs for direct determination of the 50 per cent hemolytic unit. J. Immunol., **53**, 37–50.
10. RICE, C. E. 1946. Studies of the complement-fixation reaction in virus systems. 1. Activities of vaccinia virus antigens and antisera. J. Immunol., **53**, 225–236; (a) p. 229.
11. BLISS, C. I., AND CATTELL, M. 1943. Biological Assay. Ann. Rev. Physiol., **5**, 479–539; (a) 511–518.
12. IRWIN, J. O. 1937. Statistical method applied to biological assays. J. Roy. Stat. Soc. (suppl.), **4**, 1–60; (a) 1–48; (b) 49–60; (c) 57–59.
13. IPSEN, JOHANNES. 1941. Contribution to the theory of biological standardization. N. Y. T. Nordisk, Forlag, Arnold Busck, Copenhagen, 248 p.
14. BLISS, C. I. 1938. The determination of the dosage-mortality curve from small numbers. Quart. J. Pharm., **11**, 192–216.
15. FISHER, R. A. AND YATES, F. 1938, 1943. Statistical tables for biological, agricultural and medical research, 1st and 2d ed., Edinburgh, Oliver and Boyd, Ltd.
16. GARWOOD, F. 1941. The application of maximum likelihood to dosage-mortality curves. Biometrika, **32**, 46–58.
17. WILSON, E. B. AND WORCESTER, J. 1943. The determination of L.D. 50 and its sampling error in bio-assay. 1, 2, and 3. Proc. Natl. Acad. Sci., U. S., **29**, 79–85; 114–120; 257–262.
18. WORCESTER, J. AND WILSON, E. B. 1943. A table determining L.D. 50 or the fifty per cent endpoint. Proc. Natl. Acad. Sci., U. S., **29**, 207–212.
19. BERKSON, J. 1944. Application of the logistic function to bio-assay. J. Am. Stat. Assoc., **39**, 357–365.
20. BERKSON, J. 1946. Approximation of Chi-square by "Probits" and by "Logits." J. Am. Stat. Assoc., **41**, 70–74.
21. KÄRBER, G. 1931. Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche. Arch. exptl. Path. Pharmakol., **162**, 480–483.
22. GADDUM, J. H. 1933. Reports on biological standards. 3. Methods of biological assay depending on a quantal response. Med. Research Council (Brit.), Special Rept. Series no. 183, 46 p. H. M. Stationery Office, London.
23. IRWIN, J. O. AND CHEESEMAN, E. A. 1939. On an approximate method of determining the median effective dose and its error, in the case of a quantal response. J. Hyg., **39**, 574–580.
24. IRWIN, J. O. AND CHEESEMAN, E. A. 1939. On the maximum-likelihood method of determining dosage-response curves and approximations to the median-effective dose, in cases of a quantal response. J. Roy Stat. Soc. (suppl.), **6**, 174–185.
25. REED, L. J. AND MUENCH, H. 1938. A simple method of estimating fifty per cent endpoints. Am. J. Hygiene, **27**, 493–497.
26. NEYMAN, J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J. Roy. Stat. Soc., **97**, 558–625; (a) 567–589.
27. FISHER, R. A. 1925–1946. Statistical methods for research workers. 1st–10th ed., Edinburgh, Oliver and Boyd.
28. RIDER, P. R. 1939. An introduction to modern statistical methods. New York, Wiley. 220 p.
29. RIETZ, H. L. ed. 1924. Handbook of mathematical statistics. New York, Houghton Mifflin, 221 p. (a) 58–59.
30. WILSON, E. B. AND WORCESTER, J. 1943. Bio-assay on a general curve. Proc. Natl. Acad. Sci., U. S., **29**, 150–154.

31. HUSSEY, R. G., THOMPSON, W. R., AND CALHOUN, E. T. 1927. The influence of X-rays on the development of Drosophila larvae. Science, **66**, 65–66.
32. TENNANT, R. 1931. A maximum point in an effect of prolonged X-ray irradiation upon Drosophila larvae. Science, **73**, 567–568.
33. HUSSEY, R., THOMPSON, W. R., TENNANT, R., AND CAMPBELL, N. D. 1932. The effects of radiations on biological systems. 1. Influence of high-frequency X-ray radiation upon the duration of the prepupal period of Drosophilae. J. Gen. Physiol., **16**, 207–220.
34. HUSSEY, R. AND THOMPSON, W. R. 1935. The effects of radiations on biological systems. 2. Immediate and subsequent effects of X-ray irradiation upon respiration of Drosophila larvae. J. Gen. Physiol., **18**, 669–674.
35. THOMPSON, W. R. 1935. The effects of radiations on biological systems. 3. The effect of ultraviolet light on the respiration of Drosophila larvae and the duration of their prepupal period. J. Gen. Physiol., **18**, 869–875.
36. THOMPSON, W. R. 1936. On confidence ranges for the median and other expectation distributions for populations of unknown distribution form. Ann. Math. Stat., **7**, 122–128; (a) p. 122.
37. GOODNER, K. AND HORSFALL, F. L. 1935. The protective action of Type I antipneumococcus serum in mice. I. The quantitative aspects of the mouse protection test. J. Exptl. Med., **62**, 359–374.
38. RICE, C. E. 1947. A study of the reliability of complement fixation as a method of measuring the activities of sera of high, medium, and low antibody titer. J. Immunol., **55**, 1–13.
39. CRAIG, C. C. 1942. Recent advances in mathematical statistics. Ann. Math. Stat., **13**, 74–85; (a) p. 81–82.
40. CAMP, B. H. 1931. The mathematical part of elementary statistics. D. C. Heath and Co., New York, xxi + 409 p.; (a) p. 38.
41. MALTANER, E. AND MALTANER, F. 1945. The standardization of the cardiolipin-lecithin-cholesterol antigen in the complement-fixation test for syphilis. J. Immunol., **51**, 195–214.
42. BROWN, R. 1946. The standardization of the cardiolipin-lecithin-cholesterol antigen in the precipitation test for syphilis. J. Immunol., **52**, 17–39.
43. BLISS, C. I. AND PACKARD, C. 1941. Stability of the standard dosage-effect curve for radiation. Am. J. Roentgenol. Radium Therapy, **46**, 400–404.
44. IRWIN, J. O. 1943. On the calculation of the error of biological assays. J. Hyg., **43**, 121–128.
45. BLISS, C. I. 1945. Confidence limits for biological assays. Biometrics Bull., **1**, 57–65.
46. THOMPSON, W. R. 1938. Biological applications of normal range and associated sigficance tests in ignorance of original distributior forms. Ann. Math. Stat., **9**, 281–287.
47. MALTANER, F. AND THOMPSON, W. R. 1943. Chemical analyses of the blood plasma of chicks deficient in vitamin K. Arch. Biochem., **2**, 49–54.
48. EISENHART, C. 1939. The interpretation of certain regression methods and their use in biological and industrial research. Ann. Math. Stat., **10**, 162–186.
49. HEDÉN, C.-G. 1946. On the estimation of fifty per cent end-points in serological titrimetry. J. Path. Bact., **58**, 477–81.